Article:
Li, J, Carter, LJ orcid.org/0000-0002-1146-7920 and Boxall, ABA (2020) Evaluation and development of models for estimating the sorption behaviour of pharmaceuticals in soils. Journal of Hazardous Materials, 392. 122469. ISSN 0304-3894

https://doi.org/10.1016/j.jhazmat.2020.122469

Evaluation and development of models for estimating the sorption behaviour of pharmaceuticals in soils

Jun Li,[†] Laura J. Carter,[†] and Alistair B.A. Boxall*,[†]

[†]Department of Environment and Geography, University of York, Heslington, York, YO10 5NG, UK

Abstract

Sorption is one of the key process that affects the fate and mobility of pharmaceuticals in the soil environment. Several models have been developed for estimating the sorption of organic chemicals, including ionisable compounds, in soil. However, the applicability of these models to pharmaceuticals has not been extensively tested. In this study, we generated a high-quality dataset on the sorption of twenty-one pharmaceuticals in different soil types and used these data to evaluate existing models and to develop new improved models. Sorption coefficients (Kd) of the pharmaceuticals ranged from 0.2 to 1249.2 L/kg. Existing models were unable to adequately estimate the measured sorption data. Using the data, new models were developed, incorporating molecular and soil descriptors, that outperformed the published models when evaluated against external data sets. While there is a need for further evaluation of these new models against broader sorption datasets obtained at environmentally relevant concentration, in the future they could be highly useful in supporting environmental risk assessment and prioritization efforts for pharmaceutical ingredients.

Keywords:

Ionisable compounds
Quantitative structure-property relationships
Soil properties
Environmental fate
Environmental risk assessment

1. Introduction

Pharmaceuticals are administered to prevent, diagnose and treat diseases and hence protect the health of human beings and other animals [1,2]. Following use, a large fraction of these compounds is excreted in urine and feces, which are then mostly discharged into domestic wastewater and can subsequently reach agricultural soils through irrigation using reclaimed wastewater effluent or via the application of processed or unprocessed sewage sludge to land [3,4]. A range of pharmaceuticals has been detected in agricultural soil with concentrations of antibiotics, antiepileptics, anti-inflammatory drugs, antimicrobial agents and anticoagulants being reported up to µg/kg levels [5,6].

Several studies have revealed that, following application to soil, pharmaceuticals can be taken up by soil-dwelling organisms [7-9]. The presence of pharmaceuticals in soil has been shown to reduce plant biomass and significantly affect the survival and reproduction of invertebrates [4,8]. Pharmaceutical accumulation in plants could result in exposure of humans to these compounds when they consume fruit and vegetables [3]. Furthermore, highly mobile and persistent pharmaceuticals may be transported to surface water through field runoff or leach to groundwater and subsequently affect aquatic organisms or enter human drinking water supplies [6,10,11]. Long-term exposure to pharmaceutical residues could pose a risk to ecological systems and exert adverse effects on top predators via food chain transfer [3,12].

Sorption is a key factor in determining the ultimate fate of pharmaceuticals applied to the soil environment as it influences many important processes such as the rate of leaching or the fraction of chemical that is bioavailable to organisms [13-15]. It is estimated that around 1912 pharmaceuticals are on the British market and the number is steadily increasing [16]. However, around 40 studies have been published exploring the sorption behaviour of pharmaceuticals in soil with data only being available for around 6% of the total number of pharmaceuticals and for 100 soil types. Results show that sorption coefficients for

87 pharmaceuticals in soil can vary by many orders of magnitude (e.g. 0.09 sulfameter < Kd <
88 1277873 ciprofloxacin L/kg) [17,18] and sorption coefficients for a single pharmaceutical can
89 vary by up to three orders of magnitude across different soil types (e.g. Kd values for
90 ciprofloxacin range from 726.8 to 1277873 L/kg) [17]. It is therefore clear that both chemical
91 properties and soil characteristics are important in controlling the sorption behaviour of
92 pharmaceuticals in soil [10,19-21].
93
94 Given the large number of pharmaceuticals in use and the fact that sorption data are only
95 available for a small proportion of these, to adequately understand risks of these compounds,
96 there is a need to enhance our understanding of sorption behavior. It would be cost prohibitive
97 and time-consuming to experimentally determine sorption coefficients of all pharmaceuticals
98 in the many soil types that exist in the natural environment. Modelling approaches have
99 therefore been proposed for estimating the sorption affinity of pharmaceuticals in soils. These
100 include poly-parameter Linear Free Energy Relationships and Artificial Neural Networks using
101 chemical properties alone [22,23] and models that use both chemical properties and soil
102 parameters [24-28].
103
104 Examples of models that use both chemical and soil properties include the models by Franco
105 et al. [26] and Franco and Trapp [27] who used nonlinear regression analysis to explore the
106 relationship between pharmaceutical properties and sorption behaviour in different soil
107 systems. Linear regression approaches were also proposed in the study of Kah and Brown
108 [25] and European Union technical guidance document [24] to estimate the sorption behaviour
109 of acidic organic compounds based on soil organic carbon content and pH corrected
110 lipophilicity (Log D) or hydrophobicity (Log Kow). Droge and Goss [28] developed a model that
111 estimates the sorption of bases in soil by quantifying the impact of soil organic matter, clay
112 minerals and pharmaceutical molecular structures on the contribution to sorption by both
113 hydrophobic and electrostatic interactions. Unfortunately, most of these models have been
114 developed using data published in the literature. The quality of these datasets may be

115 questionable and the spread of pharmaceuticals used to train the models may not be reflective
116 of the property distribution of all pharmaceuticals in use. There is therefore a need to evaluate
117 these models against high quality datasets on sorption behaviour of pharmaceuticals
118 representing the range of properties of pharmaceuticals in use more generally.
119
120 The aim of this study was therefore to evaluate the performance of existing models, that
121 consider the effects of both chemical and soil properties, using a high-quality dataset on
122 sorption of pharmaceuticals and, where the models are found to fail, explore the development
123 of improved models for estimating pharmaceutical sorption. The specific objectives were to: 1)
124 generate sorption data for a wide range of pharmaceuticals and soil types covering the
125 property space of pharmaceuticals more generally and soil characteristics of European
126 agricultural systems; 2) evaluate existing models against the data; and 3) use principal
127 components analysis and multi-regression methods to develop new models for pharmaceutical
128 sorption and to evaluate these against published data.
129
130 **2. Materials and methods**
131 *2.1. Study pharmaceuticals and reagents*
132
133 Twenty-one study pharmaceuticals covering thirteen therapeutic classes were purchased from
134 Sigma-Aldrich (Gillingham, UK) (purity ≥98 %). Pharmaceuticals were chosen to represent a
135 broad range of both hydrophobicity characteristics (-0.08 < Log Kow < 4.79) and ionisation
136 states at environmentally relevant pH values (-1.6 < *pKa* < 14.3). Study compounds were also
137 selected whose half-lives in soil indicated that degradation would not occur over the duration
138 of the sorption studies. Information on the physico-chemical properties, half-lives and CAS
139 number of each compound is provided in Table SI 1. HPLC grade methanol (99.9%),
140 acetonitrile (99.9 %), acetone (≥99.5 %) and water as well as calcium chloride dihydrate, and
141 potassium dihydrogen orthophosphate were obtained from Fisher Scientific (Loughborough,
142 UK). Analytical grade phosphoric acid solution (≥85 %) and formic acid (≥95 %) were

143 purchased from Sigma-Aldrich (Gillingham, UK).

144

145 *2.2. Test soils*

146

147 Five soils, covering a broad range of soil characteristics, were obtained from LandLook
148 (Midlands, UK). On receipt, the soils were air-dried and sieved through a 2-mm mesh and
149 stored in sterile sampling bags at 4 ℃ before use in the experiments. The test soils were heated
150 at 105 ℃ for 3 h to minimize biological activity prior to use. The major properties of the five
151 soils were analyzed by Forest Research Company (Surrey, UK). Detailed information on the
152 characteristics and measurement procedures of each soil is shown in Table SI 2.

153

154 *2.3. Sorption study*

155

156 Sorption studies were carried out based on OECD guideline 106 for the testing of sorption of
157 chemicals following a batch equilibrium method [29]. Preliminary sorption experiments for each
158 study compound in the test soils were conducted to identify experimental conditions for use in
159 the definitive study including the optimal soil to solution ratio, the time to reach sorption
160 equilibrium, the experimental concentration range, the appropriate test vessel, and the filtration
161 device. The optimal soil to solution ratio as well as specific concentration range of each
162 compound for each soil type were selected depending on the aqueous concentrations at
163 equilibrium and analytical method detection limits (Table SI 6). Details of the preliminary
164 sorption experiment procedures are provided in the SI Section 2.

165

166 In the definitive sorption experiments, depending on the soil and test chemical in question,
167 either 1, 2.5 or 5 g of soil (dry weight) was mixed with a specific volume of 0.01 M $CaCl_2$
168 solution (ranging from 10 to 1200 ml) to create the optimum soil to solution ratio (ranging from
169 1/1 to 1/1200, Table SI 4) in plastic or glass test vessels (selected based on stability tests for
170 two vessel types, see Table SI 4). The mixtures were shaken over 12 h in the dark to pre-

equilibrate. The soil solution mixtures were then spiked with stock solutions of the study compounds in either methanol, acetonitrile or HPLC water to give an initial concentration that ranged between 0.5 to 60 mg/L and a carrier solvent concentration of <0.1 - 0.67%. The concentration ranges of study analytes to create sorption isotherms generally differed by a factor from three to five (Table SI 4). Triplicate samples were prepared for each concentration. Control samples (containing analyte solution in 0.01 M $CaCl_2$ without soil), and one blank sample (containing $CaCl_2$ solution without study compound and soil) were prepared for each soil. All the samples were then agitated at 220 rpm in the dark at 4 ℃ for 24h or 48 h to reach sorption equilibrium (see Table SI 4). After this time, soil suspensions were centrifuged at 2500 rpm for 10 min and the resulting supernatant filtered, using 0.45 µm syringe filters, into amber glass vials for analysis.

2.4. Analytical method

Filtered samples were analysed by high performance liquid chromatography (HPLC) with diode array detection (DAD) using either a Perkin Elmer Flexar HPLC or an Agilent 1260 Infinity II HPLC instrument (The Agilent HPLC cannot be used with phosphate buffer). Separation was performed using an Agilent Zorbax Eclipse XDB C-18 column (4.6 mm × 250 mm, 5 µm pore size) at 30 ℃. The mobile phase comprised a solvent phase of either methanol or acetonitrile matched with an aqueous phase of either 0.1 % formic acid (pH= 2.7), 30 mM potassium dihydrogen orthophosphate (KH_2PO_4, pH=3.3), 25 mM potassium dihydrogen orthophosphate (KH_2PO_4, pH=3), 50 mM potassium dihydrogen orthophosphate (KH_2PO_4, pH= 4.5) or HPLC grade water adjusted to pH 2.7 with 85% phosphoric acid. The flow rate of mobile phase ranged from 0.6 to 1.4 ml min^{-1}. The injection volumes and detection wavelengths for study compounds ranged from 10 to 40 µl and 200 to 260 nm, respectively. The retention times fell within the range 2 to 4 min. Concentrations in samples were calculated based on peak area using calibration curves developed using known standards of each pharmaceutical.

The analytical methods were evaluated in terms of linearity, intra- and inter-day repeatability, matrix recovery, limit of detection (LOD) and quantitation (LOQ). The Intra-/inter-day repeatability was measured at two concentrations (2 and 20 mg/L) over 3 days. The matrix recovery was determined in supernatant samples (centrifuged from the mixture of soil and 0.01 mol/L $CaCl_2$ (1/5 and 1/200 (w/v) soil/ solution ratio)) which was then fortified with the stock solution of target pharmaceuticals at the spiking level of 5 mg/L. The limit of detection (LODs) and limits of quantification (LOQs) were calculated as three and ten times the signal-to-noise ratio, respectively [30]. Satisfactory limits of detection (0.04-0.64 mg/L) and intra-/inter-day precisions (the relative standard deviation within the range of 0-20%) were obtained for all twenty-one pharmaceuticals. With the exception of captopril, no apparent matrix interference was found for the majority of the pharmaceuticals with the average matrix recoveries of target compounds ranging from 91.25 to 103.79%. The details of the developed analytical methods and method validation results are summarised in Table SI 5 and Table SI 6.

2.5. Derivation of sorption coefficients

Linear, Freundlich and Langmuir isotherms were fitted to the data using GraphPad Prism (version 7.00). The determination of Linear, Freundlich and Langmuir isotherm constants (K_d, K_f and K_L) as well as organic carbon normalized sorption coefficient (K_{oc}) are described in the SI section 2.

2.6. Evaluation of existing predictive models

Several models, which have been proposed to predict the sorption behaviour of different classes of acidic, basic and neutral organic compounds in soil (Table 2), were evaluated using the measured sorption coefficients. The applicability and accuracy of these models were

assessed according to mathematical evidence by calculating root-mean squared deviation (RMSD) and Nash−Sutcliffe Efficiency (NSE) using the following equations (Eqs. 1, 2):

$$RMSD = \sqrt{\frac{\sum_{i=1}^{n}(Y_i^{Obs} - Y_i^{Pred})^2}{n}} \quad (1)$$

$$NSE = 1 - \left[\frac{\sum_{i=1}^{n}(Y_i^{Obs} - Y_i^{Pred})^2}{\sum_{i=1}^{n}(Y_i^{Obs} - Y^{Mean})^2}\right] \quad (2)$$

where Y_i^{Obs} and Y_i^{Pred} are the *i*th observed and predicted value, respectively. Y^{Mean} is the average of observed data and n is the number of observations. RMSD value of 0 indicates a perfect fit and less than half of the standard deviations of the observed represents a good prediction performance [31]. NSE values which can range between −∞ and 1 were used to evaluate how well the predicted values and the observed values fitted a 1:1 line. The closer that the NSE value is to 1, the better the model performance [32].

2.7. Development of new models and validation based on literature data

Principal components analysis (PCA) was performed in SPSS (version 25.0) to explore which physico-chemical properties of chemicals and soil characteristics appear to drive the sorption of each class of pharmaceuticals and to identify pharmaceutical and soil properties for use in the development of new models. The first three principal component axes were chosen to reduce the dimensionality of data according to the broken stick eigenvalue test [33].

New sorption models were then developed using 1) all soil and pharmaceuticals properties identified from the PCA; and 2) using pharmaceutical properties and soil properties, identified by the PCA, that are commonly reported in literature studies that have measured sorption of pharmaceuticals. Taking into account the degree of dissociation, multiple-linear regression analysis in the Minitab software (version 18) was used to develop new models for estimating sorption of non-ionised (neutrals, Log Kow > 0.85) and fully ionised (bases, *pKa* > 8.6) pharmaceuticals based on their molecular descriptors and soil properties. The sorption of weak

electrolytes is largely dependent on the degree of dissociation as the partitioning behaviours of dissociated and undissociated species involve different sorption mechanisms comprising different contributions to the overall sorption potential of the chemicals [26,27]. Nonlinear models were then proposed for partially ionised pharmaceuticals (weak bases, 8 > *pKa* > 4.8 and acids, 3.2 < *pKa* < 6.8) by conducting the nonlinear least squares function in the R software (R version 3.4.1). The optimum model framework applied in R software is shown in Eq.3:

$$Log\ Kd = Log(\Phi_n \cdot 10\text{^}(c_0 + c_1 \cdot X_1 + c_2 \cdot X_2 + \cdots c_i \cdot X_i) + \Phi_{ion} \cdot 10\text{^}(c_0 + c_1 \cdot X_1 + c_2 \cdot X_2 + \cdots c_i \cdot X_i)) \quad (3)$$

Where c_i and X_i represent the regression coefficients and soil and chemical parameters, respectively. Φ_n , Φ_{ion} are the neutral and ionic fractions and were derived from the Henderson-Hasselbalch equation [34].

Intercorrelated descriptors (e.g., the strong intercorrelation among hydrophobicity descriptors or the correlation between CEC and each exchangeable cation) were run separately in the regression analysis, as use of these could lead to double counting of the impact of cross-correlated parameters on the sorption.

The best performing model for each class was then identified based on 1) the number of observations used in the analysis (n), the standard error of the estimate (S), the square of the correlation coefficient (R^2), the adjusted determination coefficient (R^2_{adj}), the predicted R^2 (R^2_{pred} calculated using the leave one out approach) as well as RMSD and NSE indices; and 2) the results of an evaluation of a models predictive capability using an external evaluation data set (including 152 Kd values covering 36 pharmaceuticals) resampled from the literature (details in Table SI 10). The external evaluation dataset was also used to explore how the best performing models compared to the existing sorption models.

3. Results and discussion

3.1. Overview of sorption results

In the definitive sorption experiments, interfering peaks were observed for captopril in the UV chromatograms of the soil samples (a matrix recovery of 79.62 % was obtained at the soil/ solution ratio of 1/5), which might be attributed to the organic and inorganic components existing in the soil matrix, leading to the apparent signal suppression of the analyte response [35]. The obtained sorption coefficients of captopril were therefore not used in the evaluation of existing models and further model development. In the future, additional steps such as the use of isotopically-labeled internal standards with detection by mass spectrometry, sample dilution, or preparation of matrix-matched calibration curves are recommended to reduce the matrix effect prior to the analysis of captopril in solid samples [36].

Results of the linear, Freundlich and Langmuir isotherms fitting are presented in Table SI 7. Freundlich and linear (R^2 of 0.89 to 1.00) isotherm models better described the sorption of the pharmaceuticals, across the concentration ranges tested, than the Langmuir model (R^2 of 0.0006 to 1.00).

Sorption coefficients varied greatly within each group. Acidic pharmaceuticals exhibited lower affinity to test soils as expected, with the sorption coefficients (Kd) ranging from 0.29 L/kg (ibuprofen) to 80.45 L/kg (naproxen). For the neutral compounds, Kd values ranged from 0.20 L/kg (antipyrine) to 117.4 L/kg (disulfiram). For the bases, Kd values ranged from 0.77 L/kg (metoprolol) to 393.1 L/kg (amitriptyline). For the weak bases, values ranged from 3.24 L/kg (lamotrigine) to 1249 L/kg (perphenazine) (Table SI 7). The sorption behaviour of

307 pharmaceuticals also displayed large variability within each study soil. In soil 1, Kd values
308 ranged from 0.57 L/kg (ibuprofen) to 1181 L/kg (perphenazine). In soil 2, Kd values ranged
309 from 1.91 L/kg (captopril) to 1249 L/kg (perphenazine). In soil 3, Kd values ranged from 0.40
310 L/kg (antipyrine) to 501 L/kg (bisacodyl). In soil 4, Kd values ranged from 0.29 L/kg (ibuprofen)
311 to 861.3 L/kg (bisacodyl). Finally, in soil 5, Kd values ranged from 0.20 L/kg (antipyrine) to
312 267.4 L/kg (perphenazine) (Table SI 7). Sorption affinities of pharmaceuticals in soil 1 and 2
313 were generally higher than in the other three soils, probably due to the higher organic carbon
314 content of these soils (Figure 1). Highest variability (covering two orders of magnitudes) was
315 observed for acids among the five soils, which revealed that the soil properties (such as pH
316 and organic matter) play an important role in determining sorption behavior of acidic
317 pharmaceuticals [37].
318

319 Comparison of our findings with previous findings [10,13,18,19,23,38-43] showed that the
320 measured linear sorption coefficients of pharmaceuticals from our study for atenolol,
321 metoprolol, propranolol, amitriptyline, trimethoprim, furosemide, naproxen and carbamazepine
322 were in a similar range to sorption coefficients previously reported in the literature (Table 1).
323 For fluoxetine, our Kd values were towards the lower end of the ranges previously reported
324 and for lamotrigine, ketoprofen, ibuprofen, our Kd values were at the higher end of those
325 previously reported (Table 1). In these previous studies, a wider range of experimental
326 concentrations was typically used ranging from 0.01 µg/L to 10 mg/L which includes more
327 environmentally relevant treatments.
328

329 *3.2. Evaluation of literature models against experimental sorption data*
330

331 Ten existing models for estimating sorption of organic compounds were evaluated and
332 prediction statistics are summarized in Table 2. The best performing model overall was the

model developed by Franco and Trapp [27] for neutral pharmaceuticals which estimates sorption from the Log Kow, and which gave a RMSD of 0.409 and NSE of 0.800. Models for acids and bases performed poorly with RMSD values being greater than the standard deviation of measured sorption coefficients and negative NSEs being obtained. Moderate performance was observed for models proposed for estimating sorption of weak bases with RMSDs below standard deviation of the observations and positive NSEs being obtained. The poorer performance of models proposed for ionisable compounds is likely explained by the fact that, with the exception of the Droge and Goss model, these models consider hydrophobicity and the degree of dissociation and soil organic content and, generally, do not account for other sorption processes known to be important for ionisable compounds such as hydrogen bonding as well as electrostatic interactions (ionic exchange, charge transfer, cation bridging, ligand exchange) [10,44,45]. Therefore, in the next section, we describe work to identify key soil and pharmaceutical properties driving sorption and then move on to develop improved sorption models.

3.3. Potential factors influencing the sorption of four classes of pharmaceuticals in soil

The main factors including chemical and soil properties associated with the degree of sorption of pharmaceuticals in each class were explored by using principal components analysis (PCA) and were then used for further model development. (Details are provided in Figure 2 and Table SI 8).

3.3.1. Basic pharmaceuticals (bases, pKa > 4.8 and weak bases, 8 > *pKa* > 4.8*)*

For basic pharmaceuticals, the PCA indicated that hydrophobicity descriptors (Log Kow, Vx, Log Dow) and soil TOC had a strong positive effect on sorption and that the degree of ionisation of the pharmaceutical (F_{ion}) and soil CEC, clay and cations (Na, K, Ca) content had a weak positive effect on sorption (Table SI 8). These results suggest that bonding

361 mechanisms such as hydrophobic effects, van der Waals interactions as well as hydrogen
362 bonding interactions with organic matter, dominate the overall sorption of basic
363 pharmaceuticals in soil. Similar observations have been made in previous studies [25,46,47].
364 Moreover, most basic pharmaceuticals are predominantly in the protonated form at soil pH, so
365 some additional influence through electrostatic attraction to electronegative charged soil
366 surfaces (clay) is likely [49]. Indeed, a weak positive association of CEC and clay on sorption
367 was observed across the basic and weak basic groups that supports the existence of cation
368 exchange processes for cationic species of bases on negatively charged surfaces (clay or
369 organic matter) occupied by metal cations [10,44,49].
370
371 *3.3.2. Acidic pharmaceuticals (3.2 < pKa < 4.5)*
372
373 For acidic pharmaceuticals, the degree of dissociation (F_n) of the molecule, soil TOC and Al^{3+}
374 and Fe^{3+} had a positive effect on sorption while pH and clay content had a negative effect on
375 sorption (Table SI 8). These findings are consistent with observations from previous studies
376 where the sorption behaviour of acidic compounds was found to be strongly dependent on the
377 soil acidity [50-52]. The non-ionised species of acidic pharmaceuticals is prevalent at low pH
378 (e.g. soil 2) where the hydrophobic partitioning of neutral counterparts with organic matter via
379 van der Waals and hydrogen bonding interactions dominate the extent of sorption of acids
380 [17,45,48,51]. In addition, the strong dependence of Kd on trivalent cations suggest that cation
381 bridging between anionic form of acids and negatively charged sites and surface complexation
382 of carboxyl group to exchangeable trivalent cations on soil metal oxides and aluminosilicate
383 edge sites may be important processes for these molecules [44,46,53]. However, an
384 electrostatic repulsion interaction between the anionic form of acidic pharmaceuticals and
385 negatively charged soil surface (clay) could substantially attenuate the sorption of acids at
386 neutral and alkaline pH [10,54].
387
388 *3.3.3. Neutral pharmaceuticals (Log Kow > 0.85)*

For the neutral molecules, the PCA analysis indicated a strong positive effect of hydrophobicity and soil organic carbon on sorption (Table SI 8). This supports the hypothesis that sorption of neutral molecules is due to hydrophobic partitioning into organic matter via van der Waals and electron donor-acceptor interactions [48, 55].

3.4. Regression model development and validation

A linear regression model containing two explanatory variables (Log Kow and TOC) was generated with a good predictive capability (R^2_{pred} of 0.872) for estimating sorption coefficients for neutral pharmaceuticals (Table 3). For bases, a two-parameter model (Log Dow combined with TOC) explained 75.2% of the variation in the experimental Log Kd values. Incorporation of an additional soil property (exchangeable Na^+) into the model for bases resulted in an increase in the R^2_{pred} from 0.703 to 0.782 (Table 3). These results suggest that both hydrophobic interactions and cation exchange processes for cationic species on negatively charged surfaces occupied by metal cations drive the sorption of the basic pharmaceuticals.

Two non-linear regression models were developed for weak bases, which provided satisfactory predictive performance with the explained variance higher than 91.7% (Table 3). Molecular weight (MW) was applied to describe hydrophobic partitioning of undissociated species of weak bases, while hydrophilic factor (HF is a hydrophilicity descriptor which is calculated based on the number of carbon atoms and the number of hydrophilic groups in a molecule) was superior to other hydrophobicity descriptors in predicting the sorption of the ionic molecule species. Besides, charged surface area (simplified by the number of hydrogens bound by the charged nitrogen, Nai) and TOC were selected in explaining the sorption of ionic species, which revealed that electrostatic sorption of weak bases might be influenced by the charged surface area of the different amine types and soil organic carbon content. Furthermore, inclusion of the Ex Na^+ as model input (Model 5) yielded an improvement in the predictions of

417 Log Kd for weak bases, the R^2_{pred} increased from 0.856 to 0.892 (Table 3). The hydrophilic
418 factor (HF) combined with TOC that were found to be able to capture the variance in sorption
419 of non-ionic molecules of acids (Model 6). Molecular weight (MW) combined with soil
420 properties (CEC and soil organic carbon content) could explain the contributions of ionic
421 species to the overall sorption of acids.
422
423 The predictive performance of our developed models and existing predictive models from the
424 literature were evaluated against the literature data, which are summarised in Table 3 and
425 Table 4. Briefly, four developed models from each group all yielded good predictions ($RMSD_{test}$
426 range from 0.416 to 0.577, NSE > 0). The variability in predicted sorption coefficients by Model
427 1 agreed satisfactorily with 65 Log Kd values in the external data sets for neutral
428 pharmaceuticals across the various soil types ($RMSD_{test}$ of 0.448). In comparison, the model
429 for neutral organics proposed by Franco and Trapp [27] performed more poorly and showed
430 an underestimation of Log Kd values for hydrophobic neutrals (Log Kow > 3.36) over one order
431 of magnitude ($RMSD_{test}$ of 0.601) (see Table 4 and Figure 3). For the basic group, both the
432 proposed regression (Model 3) relying on Log Dow and TOC and the published model by
433 Franco and Trapp [27] derived from Log Kow generated the reasonable predictions and gave
434 an accuracy of a factor of 10 (N =23, Figure 3). The Model 4 proposed for weak bases
435 displayed an accurate prediction ($RMSD_{test}$ of 0.483), which outperformed the models
436 described by Franco and Trapp [27] (RMSD of 0.903 and 0.811, respectively). This revealed
437 that amine types (Nai) combined with HF provided a better estimation of the sorption of weak
438 bases compared to the single hydrophobicity descriptor (Log Kow). A satisfactory prediction
439 of sorption was feasible with Model 6 for acidic pharmaceuticals ($RMSD_{test}$ of 0.577) which
440 yielded a performance significantly superior to the two existing models proposed by Kah and
441 Brown [25] and the European Union [24] ($RMSD_{test}$ of 0.870 and 0.611, respectively), which
442 suggested that sorbate speciation is an important factor in predicting the sorption of acidic
443 pharmaceutical in soil. Similar predictions were also observed with the models developed by
444 Franco et al. [26] and Franco and Trapp [27], with the average errors of 0.558 and 0.573,

445 respectively.

446

447 Overall, the model evaluation results based on the independent data set demonstrates that
448 the sorption affinity of the partially ionised pharmaceuticals could be estimated accurately by
449 weighting the contributions of neutral and ionic molecule species separately. The multiple-
450 linear regression models to estimate the sorption coefficient of the nonionised and fully ionised
451 pharmaceuticals yielded appropriate predictions by incorporating molecular and soil properties
452 (all predicted Log Kd values within a factor of 10). However, the better Models 2 and 5 for basic
453 and weak basic pharmaceuticals and sorption model developed by Droge and Goss (2013)
454 [26] containing the soil descriptors (exchangeable Na^+ and CEC) could not be evaluated due
455 to the incomplete record of soil properties being reported in many studies in the literature. The
456 predictive performance of these models is worthy of further validation through the generation
457 of additional experimental data on a wider range of pharmaceuticals and soil types and
458 employing more environmentally-relevant concentrations.

459

460 **4. Conclusion**

461 In this study, the sorption behaviour of twenty-one pharmaceuticals across thirteen therapeutic
462 classes was investigated in five test soils with different properties. Use of the data to evaluate
463 existing sorption models, relying solely on Log Kow, for estimating sorption of neutral
464 pharmaceuticals indicated that these models worked well. However, comparison of the
465 sorption coefficients, obtained in the experiments, with predictions from existing models for
466 estimating sorption of ionisable compounds showed that the models performed poorly for
467 pharmaceuticals. Work was therefore done to develop new modelling approaches. An initial
468 PCA analysis indicated that the sorption of the study pharmaceuticals was driven by
469 hydrophobic forces as well as electrostatic interactions and a range of soil parameters. Using
470 this knowledge, new models were developed for estimating sorption coefficients for
471 pharmaceuticals. Evaluation of these new models against an independent dataset obtained
472 from the literature showed that the models were on par with (model for bases and acids) or

473 superior to (model for neutrals and weak bases) existing models.

474

475 While our study was more extensive than previous investigations of this type in terms of the
476 range of pharmaceuticals and soil investigated, it still only focused on a subset of the
477 pharmaceuticals in a small number of soils. The study also employed concentrations greater
478 than concentrations typically observed in the environment. In the future, we recommend that
479 further work is done at lower concentrations that are environmentally relevant and using a
480 wider concentration range to further evaluate the models and, if appropriate, further refine the
481 relationships. These models would allow us to predict sorption behavior of
482 pharmaceuticals under realistic environmental conditions and could be invaluable for not only
483 characterizing the environmental risks of pharmaceuticals in soil environments but also in
484 sediment-water systems.

485

486 **Acknowledgments**

487 We would like to thank Matt Pickering for his valuable comments in HPLC-method
488 development. We are grateful to one anonymous reviewer for their detailed and constructive
489 comments on an earlier version of this manuscript.

490

491 **Supporting information description**

492 Detailed information on study pharmaceuticals and soils, the preliminary experiment
493 procedures and analytical methods, sorption isotherms for study pharmaceuticals, results of
494 principle component analysis, goodness of fit of developed models and existing predictive
495 models against the external data sets as well as details of external evaluation data sets.

496

497 **Author information**

498 **Corresponding Author**

499 *E-mail: alistair.boxall@york.ac.uk; Tel: +44 (0)1904 324791; fax: +44 (0)1904 322998.

Funding

The study was performed as part of the *Intelligence Led Assessment of Pharmaceuticals in the Environment* project (iPiE Grant Number: 115735), which was funded by the EU/EFPiA Innovative Medicines Initiative Joint Undertaking.

References

1. Boxall, A. B., Kolpin, D. W., Halling-Sørensen, B., & Tolls, J. Peer reviewed: are veterinary medicines causing environmental risks? *Environmental science & technology*, **2003**, *37*(15), 286A-294A.

2. Li, W. C. Occurrence, sources, and fate of pharmaceuticals in aquatic environment and soil. *Environmental pollution*, **2014**, *187*, 193-201.

3. Shenker, M., Harush, D., Ben-Ari, J., & Chefetz, B. Uptake of carbamazepine by cucumber plants–A case study related to irrigation with reclaimed wastewater. *Chemosphere*, **2011**, *82*(6), 905-910.

4. Carter, L. J., Williams, M., Böttcher, C., & Kookana, R. S. Uptake of pharmaceuticals influences plant development and affects nutrient and hormone homeostases. *Environmental science & technology*, **2015**, *49*(20), 12509-12518.

5. Ho, Y. B., Zakaria, M. P., Latif, P. A., & Saari, N. Occurrence of veterinary antibiotics and progesterone in broiler manure and agricultural soil in Malaysia. *science of the Total Environment*, **2014**, *488*, 261-267.

6. Qin, Q., Chen, X., & Zhuang, J. The fate and impact of pharmaceuticals and personal care products in agricultural soils irrigated with reclaimed water. *Critical Reviews in Environmental Science and Technology*, **2015**, *45*(13), 1379-1408.

7. Carter, L. J., Garman, C. D., Ryan, J., Dowle, A., Bergström, E., Thomas-Oates, J., & Boxall, A. B. Fate and uptake of pharmaceuticals in soil–earthworm systems. *Environmental science & technology*, **2014**, *48*(10), 5955-5963.

8. Kinney, C. A., Campbell, B. R., Thompson, R., Furlong, E. T., Kolpin, D. W., Burkhardt, M. R., ... & Hay, A. G. Earthworm bioassays and seedling emergence for monitoring toxicity, aging and bioaccumulation of anthropogenic waste indicator compounds in biosolids–amended soil. *Science of the total environment*, **2012**, *433*, 507-515.

9. Pan, M., Wong, C. K., & Chu, L. M. Distribution of antibiotics in wastewater-irrigated soils and their accumulation in vegetable crops in the Pearl River Delta, Southern China. *Journal of agricultural and food chemistry*, **2014**, *62*(46), 11062-11069.

10. Kodešová, R., Grabic, R., Kočárek, M., Klement, A., Golovko, O., Fér, M., ... & Jakšík, O. Pharmaceuticals' sorptions relative to properties of thirteen different soils. *Science of the total environment*, **2015**, *511*, 435-443.

11. Tolls, J. Sorption of veterinary pharmaceuticals in soils: a review. *Environmental science & technology*, **2001**, *35*(17), 3397-3406.

12. Wu, X., Dodgen, L. K., Conkle, J. L., & Gan, J. Plant uptake of pharmaceutical and personal care products from recycled water and biosolids: a review. *Science of the Total Environment*, **2015**, *536*, 655-666.

13. Drillia, P., Stamatelatou, K., & Lyberatos, G. Fate and mobility of pharmaceuticals in solid matrices. *Chemosphere*, **2005**, *60*(8), 1034-1044.

14. Wang, S., & Wang, H. Adsorption behavior of antibiotic in soil environment: a critical review. *Frontiers of Environmental Science & Engineering*, **2015**, *9*(4), 565-574.

15. Carter, L. J., Ryan, J. J., & Boxall, A. B. Effects of soil properties on the uptake of pharmaceuticals into earthworms. *Environmental pollution*, **2016**, *213*, 922-931.

16. Burns, E. E., Carter, L. J., Snape, J., Thomas-Oates, J., & Boxall, A. B. Application of prioritization approaches to optimize environmental monitoring and testing of pharmaceuticals. *Journal of Toxicology and Environmental Health, Part B*, **2018**, *21*(3), 115-141.

17. Leal, R. M. P., Alleoni, L. R. F., Tornisielo, V. L., & Regitano, J. B. Sorption of fluoroquinolones and sulfonamides in 13 Brazilian soils. *Chemosphere*, **2013**, *92*(8), 979-985.

18. Zhang, Y. L., Lin, S. S., Dai, C. M., Shi, L., & Zhou, X. F. Sorption–desorption and transport of trimethoprim and sulfonamide antibiotics in agricultural soil: effect of soil type, dissolved organic matter, and pH. *Environmental Science and Pollution Research*, **2014**, *21*(9), 5827-5835.

19. Williams, M., Ong, P. L., Williams, D. B., & Kookana, R. S. Estimating the sorption of pharmaceuticals based on their pharmacological distribution. *Environmental toxicology and chemistry*, **2009**, *28*(12), 2572-2579.

20. Kim, Y., Lim, S., Han, M., & Cho, J. Sorption characteristics of oxytetracycline, amoxicillin, and sulfathiazole in two different soil types. *Geoderma*, **2012**, *185*, 97-101.

21. Pan, M., & Chu, L. M. Adsorption and degradation of five selected antibiotics in agricultural soil. *Science of the Total Environment*, **2016**, *545*, 48-56.

22. Bronner, G., & Goss, K. U. Predicting sorption of pesticides and other multifunctional organic chemicals to soil organic carbon. *Environmental science & technology*, **2010**, *45*(4), 1313-1319.

23. Barron, L., Havel, J., Purcell, M., Szpak, M., Kelleher, B., & Paull, B. (2009). Predicting sorption of pharmaceuticals and personal care products onto soil and digested sludge using artificial neural networks. *Analyst*, **2009**, *134*(4), 663-670.

24. European Union. Technical Guidance Document (TGD) on Risk Assessment of Chemical Substances following European Regulations and Directives, Parts III. *Technical Report Number EUR 20418 EN/1-4*, **2003**.

25. Kah, M., & Brown, C. D. Prediction of the adsorption of ionizable pesticides in soils. *Journal of agricultural and food chemistry*, **2007**, *55*(6), 2312-2322.

26. Franco, A., Fu, W., & Trapp, S. Influence of soil pH on the sorption of ionizable chemicals: modeling advances. *Environmental Toxicology and Chemistry*, **2009**, *28*(3), 458-464.

27. Franco A., & Trapp S. Estimation of the soil-water partition coefficient normalized to organic carbon for ionisable organic chemicals. *Environmental Toxicology and Chemistry*, **2008**, *27*(10): 1995–2004.

608 28. Droge, S. T., & Goss, K. U. Development and evaluation of a new sorption model for
609 organic cations in soil: contributions from organic matter and clay minerals. *Environmental*
610 *science & technology*, **2013**, *47*(24), 14233-14241.
611
612 29. OECD Guidelines for the Testing of Chemicals: Test No. 106 Adsorption Desorption Using
613 a Batch Equilibrium Method; Organization for Economic Cooperation and Development: Paris,
614 France, **2000**. Http: www.oecd.org/env/ehs/testing/TG_List_EN_Jul_2013.pdf.
615
616 30. Doretto, K. M., & Rath, S. Sorption of sulfadiazine on Brazilian soils. *Chemosphere*, **2013**,
617 *90*(6), 2027-2034.
618
619 31. Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., & Veith, T. L.
620 Model evaluation guidelines for systematic quantification of accuracy in watershed
621 simulations. *Transactions of the ASABE*, **2007**, *50*(3), 885-900.
622
623 32. Singh, J., Knapp, H. V., Arnold, J. G., & Demissie, M. HydroLogical modeling of the iroquois
624 river watershed using HSPF and SWAT 1. *JAWRA Journal of the American Water Resources*
625 *Association*, **2005**, *41*(2), 343-360.
626
627 33. Legendre P, Legendre L. *Numerical EcoLogy*. Elsevier Science, Amsterdam, The
628 Netherlands,**1998.**
629
630 34. Henderson, L. J. Concerning the relationship between the strength of acids and their
631 capacity to preserve neutrality. *American Journal of Physiology-Legacy Content*, **1908**, *21*(2),
632 173-179.
633
634 35. Yu, K., Li, B., & Zhang, T. Direct rapid analysis of multiple PPCPs in municipal wastewater
635 using ultrahigh performance liquid chromatography–tandem mass spectrometry without SPE
636 pre-concentration. *Analytica chimica acta*, **2012**, *738*, 59-68.
637
638 36. Campos-Mañas, M. C., Plaza-Bolaños, P., Sánchez-Pérez, J. A., Malato, S., & Agüera, A.
639 Fast determination of pesticides and other contaminants of emerging concern in treated
640 wastewater using direct injection coupled to highly sensitive ultra-high performance liquid
641 chromatography-tandem mass spectrometry. *Journal of Chromatography A*, **2017**, *1507*, 84-
642 94.
643

644 37. Tülp, H. C., Fenner, K., Schwarzenbach, R. P., & Goss, K. U. pH-dependent sorption of
645 acidic organic chemicals to soil organic matter. *Environmental science & technology*, **2009**,
646 *43*(24), 9189-9195.
647
648 38. Monteiro, S. *Fate of human-use pharmaceuticals in the soil environment*, **2008**, (Doctoral
649 dissertation, York).
650
651 39. Xu, J., Chen, W., Wu, L., & Chang, A. C. Adsorption and degradation of ketoprofen in
652 soils. *Journal of environmental quality*, **2009**, *38*(3), 1177-1182.
653
654 40. Xu, J., Wu, L., & Chang, A. C. Degradation and adsorption of selected pharmaceuticals
655 and personal care products (PPCPs) in agricultural soils. *Chemosphere*, **2009**, *77*(10), 1299-
656 1305.
657
658 41. Paz, A., Tadmor, G., Malchi, T., Blotevogel, J., Borch, T., Polubesova, T., & Chefetz, B. Fate
659 of carbamazepine, its metabolites, and lamotrigine in soils irrigated with reclaimed wastewater:
660 Sorption, leaching and plant uptake. *Chemosphere*, **2016**, *160*, 22-29.
661
662 42. Durán–Álvarez, J. C., Prado, B., Ferroud, A., Juayerk, N., & Jiménez-Cisneros, B. Sorption,
663 desorption and displacement of ibuprofen, estrone, and 17β estradiol in wastewater irrigated
664 and rainfed agricultural soils. *Science of the Total Environment*, **2014**, *473*, 189-198.
665
666 43. Lin, K., & Gan, J. Sorption and degradation of wastewater-associated non-steroidal anti-
667 inflammatory drugs and antibiotics in soils. *Chemosphere*, **2011**, *83*(3), 240-246.
668
669 44. Vasudevan, D., Bruland, G. L., Torrance, B. S., Upchurch, V. G., & MacKay, A. A. pH-
670 dependent ciprofloxacin sorption to soils: Interaction mechanisms and soil factors influencing
671 sorption. *Geoderma*, **2009**, *151*(3), 68-76.
672
673 45. Zhang, Y., Price, G. W., Jamieson, R., Burton, D., & Khosravi, K. Sorption and desorption
674 of selected non-steroidal anti-inflammatory drugs in an agricultural loam-textured
675 soil. *Chemosphere*, **2017**, *174*, 628-637.
676
677 46. Kah, M., & Brown, C. D. Adsorption of ionisable pesticides in soils. In *Reviews of*
678 *environmental contamination and toxicology*. Springer New York, **2006**, 149-217.
679

680 47. Al-Khazrajy, O. S., & Boxall, A. B. Impacts of compound properties and sediment
681 characteristics on the sorption behaviour of pharmaceuticals in aquatic systems. *Journal of*
682 *hazardous materials*, **2016**, *317*, 198-209.
683
684 48. Klement, A., Kodešová, R., Bauerová, M., Golovko, O., Kočárek, M., Fér, M., ... & Grabic,
685 R. Sorption of citalopram, irbesartan and fexofenadine in soils: Estimation of sorption
686 coefficients from soil properties. *Chemosphere*, **2018**, *195*, 615-623.
687
688 49. Hyland, K. C., Dickenson, E. R., Drewes, J. E., & Higgins, C. P. Sorption of ionized and
689 neutral emerging trace organic compounds onto activated sludge from different wastewater
690 treatment configurations. *Water research*, **2012**, *46*(6), 1958-1968.
691
692 50. Chefetz, B., Mualem, T., & Ben-Ari, J. Sorption and mobility of pharmaceutical compounds
693 in soil irrigated with reclaimed wastewater. *Chemosphere*, **2008**, *73*(8), 1335-1343.
694
695 51. Revitt, D. M., BaLogh, T., & Jones, H. Sorption behaviours and transport potentials for
696 selected pharmaceuticals and triclosan in two sterilised soils. *Journal of soils and*
697 *sediments*, **2015**, *15*(3), 594-606.
698
699 52. Foolad, M., Hu, J., Tran, N. H., & Ong, S. L. Sorption and biodegradation characteristics
700 of the selected pharmaceuticals and personal care products onto tropical soil. *Water Science*
701 *and Technology*, **2016**, *73*(1), 51-59.
702
703 53. Bui, T. X., & Choi, H. Influence of ionic strength, anions, cations, and natural organic matter
704 on the adsorption of pharmaceuticals to silica. *Chemosphere*, **2010**, *80*(7), 681-686.
705
706 54. Maoz, A., & Chefetz, B. Sorption of the pharmaceuticals carbamazepine and naproxen to
707 dissolved organic matter: role of structural fractions. *Water research*, **2010**, *44*(3), 981-989.
708
709 55. Williams, C. F., & Adamsen, F. J. Sorption–desorption of carbamazepine from irrigated
710 soils. *Journal of environmental quality*, **2006**, *35*(5), 1779-1783.

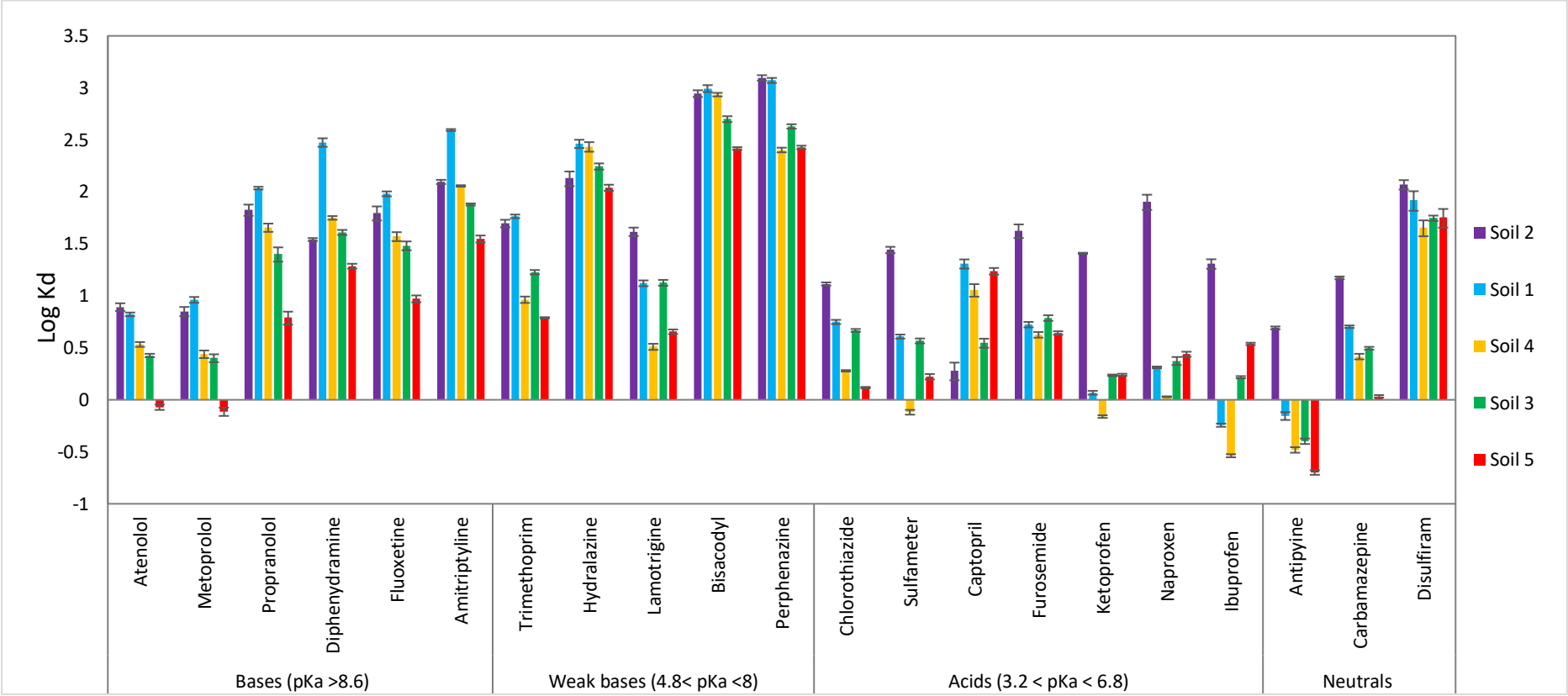


Figure 1. Logarithm of the linear sorption coefficient (Log Kd values) (±SE) for all the investigated pharmaceuticals in the five study soils. Compounds within a group ordered from low to high Log Kow. Soil organic carbon content increased in the order of soil 2 > soil 1 > soil 4 > soil 3 > soil 5.

Table 1. Comparison of the sorption coefficient (Kd) measured in present study and reported Kd values of pharmaceuticals in soil environments.

Compound	Measured	Literature
	Kd (L/kg)	Kd (L/kg) (Reference)
Atenolol	0.85-7.81	1.61-7.08 (19); 15 (23); 1.88-4.8 (10)
Metoprolol	0.77-9.16	25.4-75 (19); 20 (23); 1.36-3.83 (10)
Propranolol	6.16-108.7	58 (23); 16.3-199 (13)
Diphenhydramine	19.3-299.2	n.d.
Fluoxetine	9.38-95.78	146-234.8 (38)
Amitriptyline	35.29-393.1	138 (23)
Trimethoprim	6.15-58.16	4.67-109(19); 26 (23); 1.16 (10); 7.06-9.21 (18); 7.42 (43)
Hydralazine	109.70-290.36	n.d.
Lamotrigine	3.24-41.45	0.73-2.64 (41)
Bisacodyl	261.1-986.2	n.d.
Perphenazine	252.9-1249	n.d.
Chlorothiazide	1.31-13	n.d.
Sulfameter	0.76-27.65	0.09-0.17 (18)
Captopril	1.91-20.34	n.d.
Furosemide	4.22-42.3	27 (23)
Ketoprofen	0.69-25.59	0.09-9.59 (19); 9 (23); 1.26-8.24 (39)
Naproxen	1.07-80.45	0.23-17.5 (19); 11(23); 10.1-252.9 (38); 1.24-16.49 (40); 2.39-4.41 (12)
Ibuprofen	0.29-20.32	0.15-3.01(19); 21 (23); 0.56-3.71(40); 1.18(42); 1.08-1.14 (43)
Antipyrine	0.20-4.92	n.d.
Carbamazepine	1.08-14.88	0.53-16.7(19); 13 (23); 0.43 (10); 0.49-37 (13); 4.7-32.8 (38); 0.53-1.25 (41)
Disulfiram	45.28-117.4	n.d.

n.d.: no data.

Table 2. Evaluation of existing regression models for estimating the sorption behaviour of neutral, basic and acidic organic compounds in soil (The predicted organic carbon-normalised sorption coefficients (Log Koc) were converted to Log Kd to allow comparison to experimental data).

Class	Regression model		N	R^2	SD	RMSD	NSE
Neutrals	Franco and Trapp (2008)	$Log\ Koc = 0.5 * Log\ P + 1.13$	N=15	0.907	0.947	0.409	0.800
Bases	Droge and Goss (2013)	$Kd = K_{CEC,Clays}(CEC_{Soil} - 3.4 f_{oc}) + f_{oc} * D_{oc,IE}$	N=25	0.091	0.745	1.311	-2.230
	Franco and Trapp (2008) base model A	$Log\ Koc = Log\ (\phi n * 10^{0.21 * Log\ P + 2.24} + \phi ion * 10^{0.42 * Log\ P + 2.19})$	N=30	0.709	0.710	0.780	-0.247
	Franco and Trapp (2008) base model B	$Log\ Koc = Log\ (\phi n * 10^{0.37 * Log\ P + 1.7} + \phi ion * 10^{pKa^{0.65} * f^{0.14}})$	N=30	0.529	0.710	1.077	-1.376
Weak Bases	Franco and Trapp (2008) base model A	$Log\ Koc = Log\ (\phi n * 10^{0.21 * Log\ P + 2.24} + \phi ion * 10^{0.42 * Log\ P + 2.19})$	N=25	0.473	0.816	0.691	0.253
	Franco and Trapp. (2008) base model B	$Log\ Koc = Log\ (\phi n * 10^{0.37 * Log\ P + 1.7} + \phi ion * 10^{pKa^{0.65} * f^{0.14}})$	N=25	0.309	0.816	0.686	0.263
Acids	Franco and Trapp (2008)	$Log\ Koc = Log\ (\phi n * 10^{0.54 * Log\ P + 1.11} + \phi ion * 10^{0.11 * Log\ P + 1.54})$	N=30	0.166	0.576	0.640	-0.276
	Franco et al. (2009)	$Koc = \frac{10^{0.54 * Log\ P + 1.11}}{1 + 10^{(pH - 0.6 - pKa)}} + \frac{10^{0.11 * Log\ P + 1.54}}{1 + 10^{(pKa - pH + 0.6)}}$	N=30	0.115	0.576	0.694	-0.503
	Kah and Brown (2007)	$Log\ Kd = 0.13 * Log\ D + 1.02\ Log\ OC\ - 1.51$	N=30	0.282	0.576	0.655	-3.359
	European Union (2003)	$Log\ Koc = 0.6 * Log\ P + 0.32$	N=30	0.001	0.576	1.127	-2.961

f_{oc}: fraction organic carbon in soil;

Log P: the octanol–water partition coefficient;

pKa: acid-dissociation coefficient;

ϕn, ϕion: fraction of neutral and ionic species;

f: fraction of compound in the lipophilic phase, f = Kow/(Kow+1);

Log D: lipophilicity corrected to soil pH;

$K_{CEC,Clay}$ and $D_{OC,IE}$ are CEC-normalized and soil organic matter-normalized sorption coefficients, respectively. Log $K_{CEC,Clay}$ = 1.22 Vx - 0.22Nai + 1.09; Log $D_{oc,IE}$ = 1.53Vx + 0.32Nai – 0.27;

Vx: molecular volume was determined following the approach described in Abraham and McGowan's, (1987);

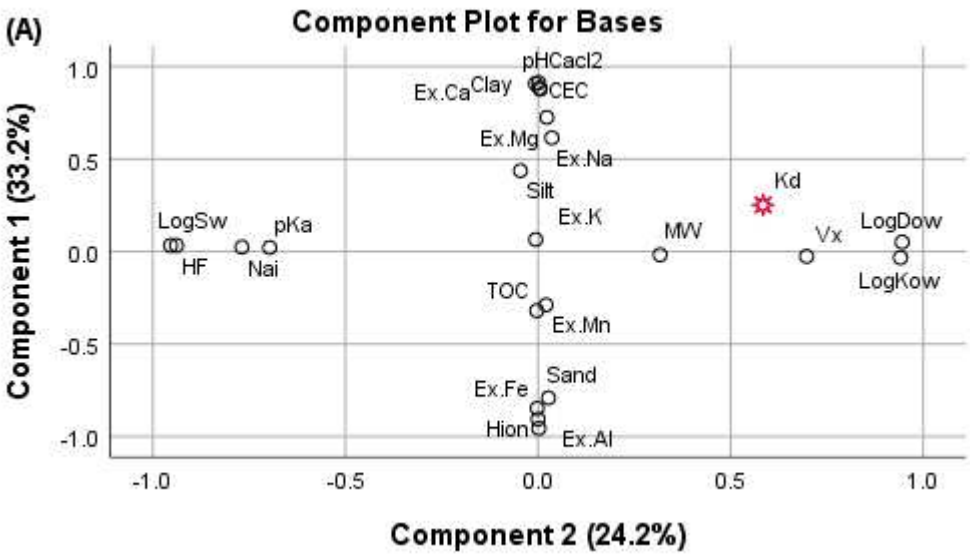
Nai: number of hydrogens bound by the charged nitrogen;

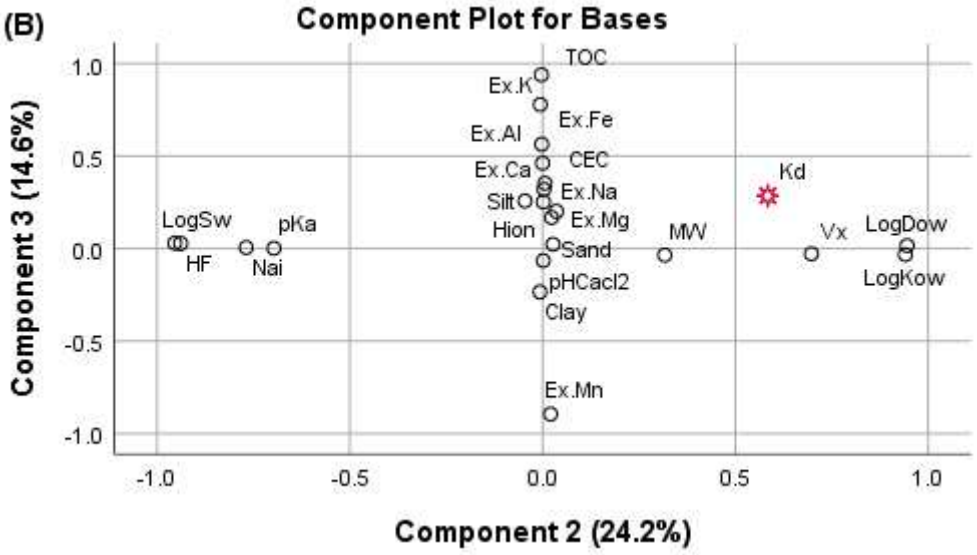
N: Number of observations;

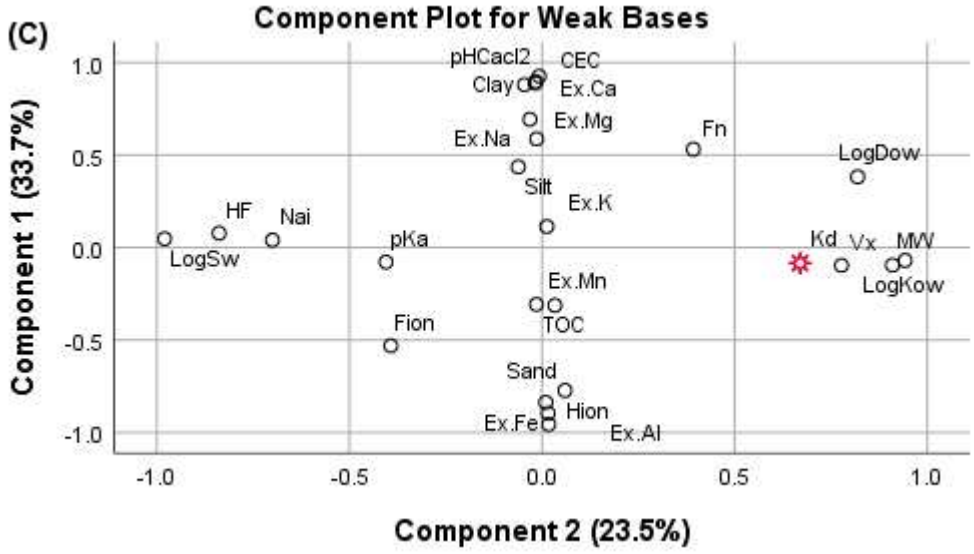
SD: Standard deviation of the observation;

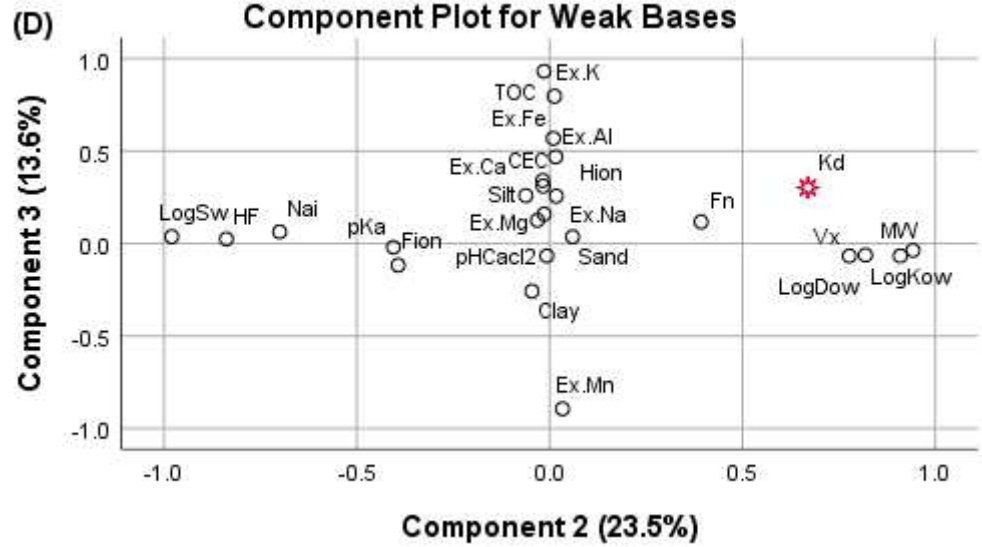
RMSD: Root mean square deviation;

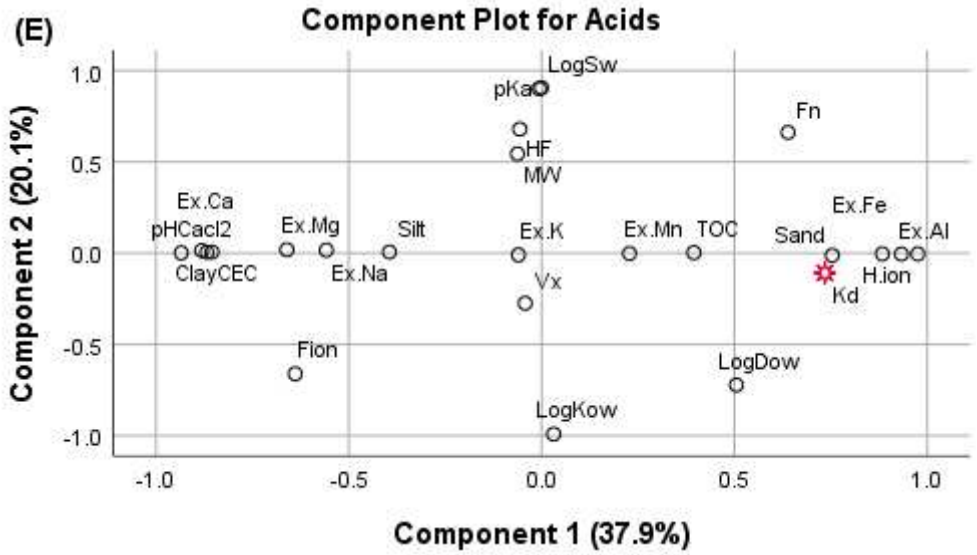
NSE: Nash-Sutcliffe Efficiency.

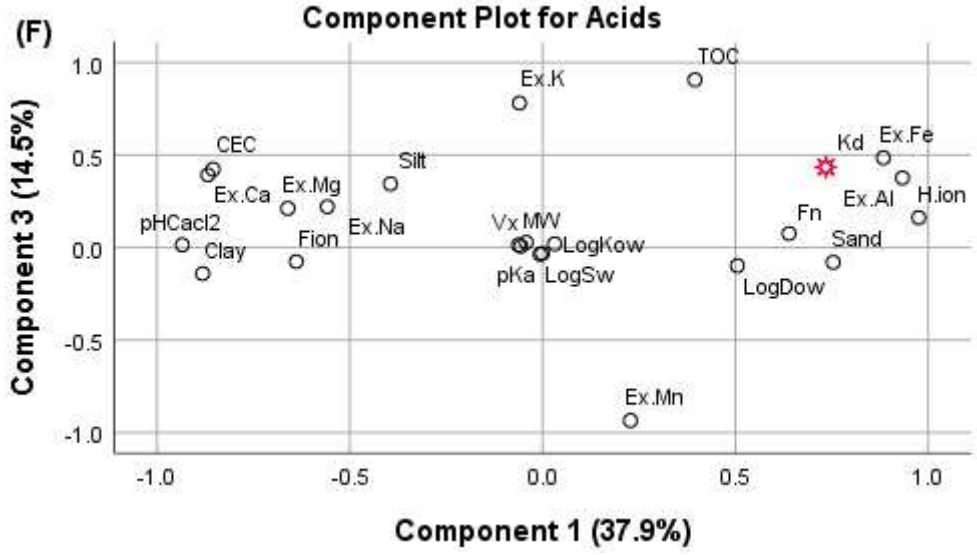

(A)
Component Plot for Bases
Component 1 (33.2%)
Component 2 (24.2%)
pHCacl2
Ex.Ca
Clay
CEC
Ex.Mg
Ex.Na
Silt
Kd
LogSw
pKa
Ex.K
MW
Vx
LogDow
HF
Nai
LogKow
TOC
Ex.Mn
Sand
Ex.Fe
Hion
Ex.Al

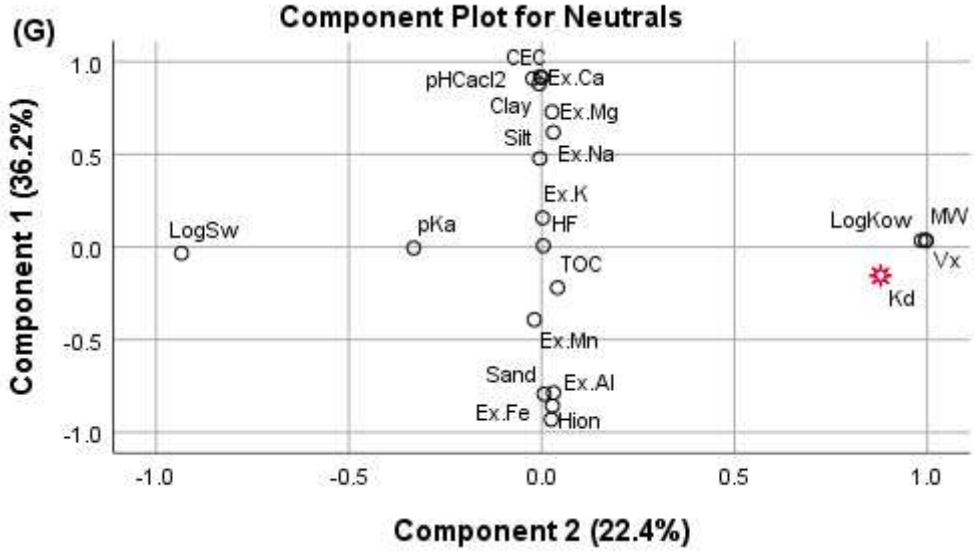

(B)
Component Plot for Bases
Component 3 (14.6%)
Component 2 (24.2%)
TOC
Ex.K
Ex.Fe
Ex.Al
Ex.Ca
CEC
Kd
Silt
Ex.Na
LogSw
pKa
Hion
Ex.Mg
MW
Vx
LogDow
HF
Nai
Sand
LogKow
pHCacl2
Clay
Ex.Mn


(C)
Component Plot for Weak Bases
Component 1 (33.7%)
Component 2 (23.5%)
pHCacl2
CEC
Clay
Ex.Ca
Ex.Mg
Fn
Ex.Na
LogDow
Silt
Ex.K
HF
Nai
pKa
Kd
Vx
MW
LogSw
Ex.Mn
LogKow
Fion
TOC
Sand
Hion
Ex.Fe
Ex.Al


(D)
Component Plot for Weak Bases
Component 3 (13.6%)
Component 2 (23.5%)
Ex.K
TOC
Ex.Fe
Ex.Al
Ex.Ca
CEC
Hion
Kd
Silt
Fn
LogSw
HF
Nai
Ex.Mg
Ex.Na
pKa
Fion
Vx
MW
pHCacl2
Sand
LogDow
LogKow
Clay
Ex.Mn


(E)
Component Plot for Acids
LogSw
pKa
Fn
HF
MW
Ex.Ca
pHCacl2
Ex.Mg
Silt
Ex.K
Ex.Mn
TOC
Sand
Ex.Fe
Ex.Al
ClayCEC
Ex.Na
Vx
H.ion
Kd
Fion
LogDow
LogKow
Component 2 (20.1%)
Component 1 (37.9%)
1.0
0.5
0.0
-0.5
-1.0


(F)
Component Plot for Acids
TOC
Ex.K
CEC
Kd
Ex.Fe
Silt
Ex.Ca
Ex.Mg
Ex.Al
H.ion
pHCacl2
Vx
MW
Fn
Ex.Na
Clay
Fion
LogKow
Sand
pKa
LogSw
LogDow
Ex.Mn
Component 3 (14.5%)
Component 1 (37.9%)
1.0
0.5
0.0
-0.5
-1.0


(G)
Component Plot for Neutrals
CEC
pHCacl2
Ex.Ca
Clay
Ex.Mg
Silt
Ex.Na
Ex.K
LogSw
pKa
HF
LogKow
MW
Vx
TOC
Kd
Ex.Mn
Sand
Ex.Al
Ex.Fe
Hion
Component 1 (36.2%)
Component 2 (22.4%)
1.0
0.5
0.0
-0.5
-1.0

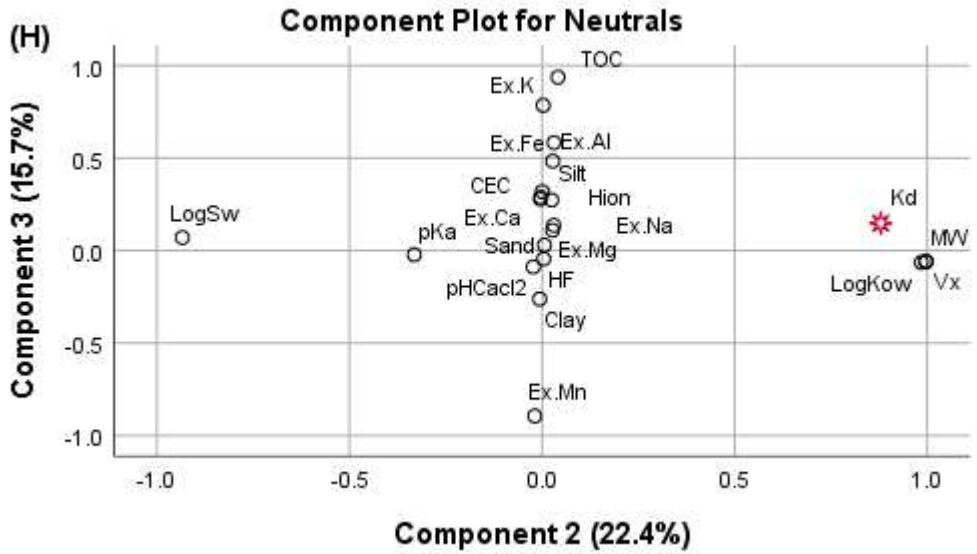

(H)
Component Plot for Neutrals
TOC
Ex.K
Ex.Fe
Ex.Al
Silt
CEC
Hion
Kd
LogSw
Ex.Ca
pKa
Ex.Na
Sand
Ex.Mg
MW
LogKow
Vx
HF
pHCacl2
Clay
Ex.Mn
Component 3 (15.7%)
Component 2 (22.4%)
1.0
0.5
0.0
-0.5
-1.0

Figure 2. Principal component analysis loading plots for Kd, soil and pharmaceutical properties for basic compounds (A,B); weak basic compounds (C,D); acidic compounds (E,F); and for neutral compounds (G,H).

Table 3. Multiple linear and non-linear regression equations for predicting sorption coefficients of pharmaceuticals in soils

Class	Model	Equation	Training						Test				
			N	SE	R^2	R^2_{adj}	R^2_{pred}	$RMSD_{train}$	N	SD	R^2_{test}	$RMSD_{test}$	NSE
Neutrals (Log Kow > 0.85)	1	$Log\ Kd = 0.779 * Log\ Kow + 0.211 * TOC - 1.729$	15	0.265	0.933	0.921	0.872	0.237	65	0.637	0.543	0.448	0.497
Bases (*pKa* > 8)	2	$Log\ Kd = 0.312 * Log\ Dow + 0.171 * TOC + 4.164 * ExNa + 0.336$	30	0.306	0.834	0.815	0.782	0.284	n.d.				
	3	$Log\ Kd = 0.315 * Log\ Dow + 0.188 * TOC + 0.585$	30	0.367	0.752	0.733	0.703	0.348	23	0.447	0.721	0.416	0.094
Weak bases (*pKa* < 8)	4	$Log\ Kd = Log\ (\phi n * 10^{0.021*MW-4.7} + \phi ion * 10^{-0.535*HF+0.345*Nai+0.145*TOC+1.559})$	25	0.264	0.917	0.895	0.856	0.230	20	1.082	0.816	0.483	0.790
	5	$Log\ Kd = Log\ (\phi n * 10^{0.021*MW-4.979} + \phi ion * 10^{-0.54*HF+0.331*Nai+3.208*Ex\ Na+0.139*TOC+1.389})$	25	0.228	0.942	0.922	0.892	0.193	n.d.				
Acids (6.8 > *pKa* > 3.2)	6	$Log\ Kd = Log\ (\phi n * 10^{-0.313*HF+0.191*TOC+0.417} + \phi ion * 10^{0.0083*MW-0.038*CEC+0.301*TOC-2.36})$	30	0.198	0.906	0.886	0.842	0.174	44	0.733	0.456	0.577	0.366

All the regression descriptors were statistically significant at the 0.05 level.

Log Kow, *pKa*, MW, Log Dow are the partition coefficient of the neutral molecule, dissociation constant, molecular weight, pH-dependent octanol-water distribution coefficient , respectively, which were calculated by the software ACD/Labs(http://ilab.cds.rsc.org/). HF (hydrophilic factor) was obtained from alvaDesc (v1.0.8).

ϕn, ϕion are the fraction of neutral and ionic species, respectively.

Nai: number of hydrogens bound by the charged nitrogen;

Ex Na^+ and CEC are exchangeable sodium and cation exchange capacity (cmol+/kg), respectively. Clay and TOC are clay content and total organic carbon content (%) in soil, respectively.

N_{train}, N_{test} are the number of the experimental sorption coefficients and published sorption coefficients, respectively.

SE, SD_{test} are the standard error of the fitted model and standard deviation of published sorption coefficients.

R^2_{adj}, R^2_{pred} is the adjusted R^2, predicted R^2 of developed models.

$RMSD_{train}$, $RMSD_{test}$ are root mean square deviation of experimental data against predicted data and test data against predicted data, respectively.

NSE is the Nash−Sutcliffe Efficiency value.

n.d.: no data.

Table 4. Predictive performance of existing models against literature data.

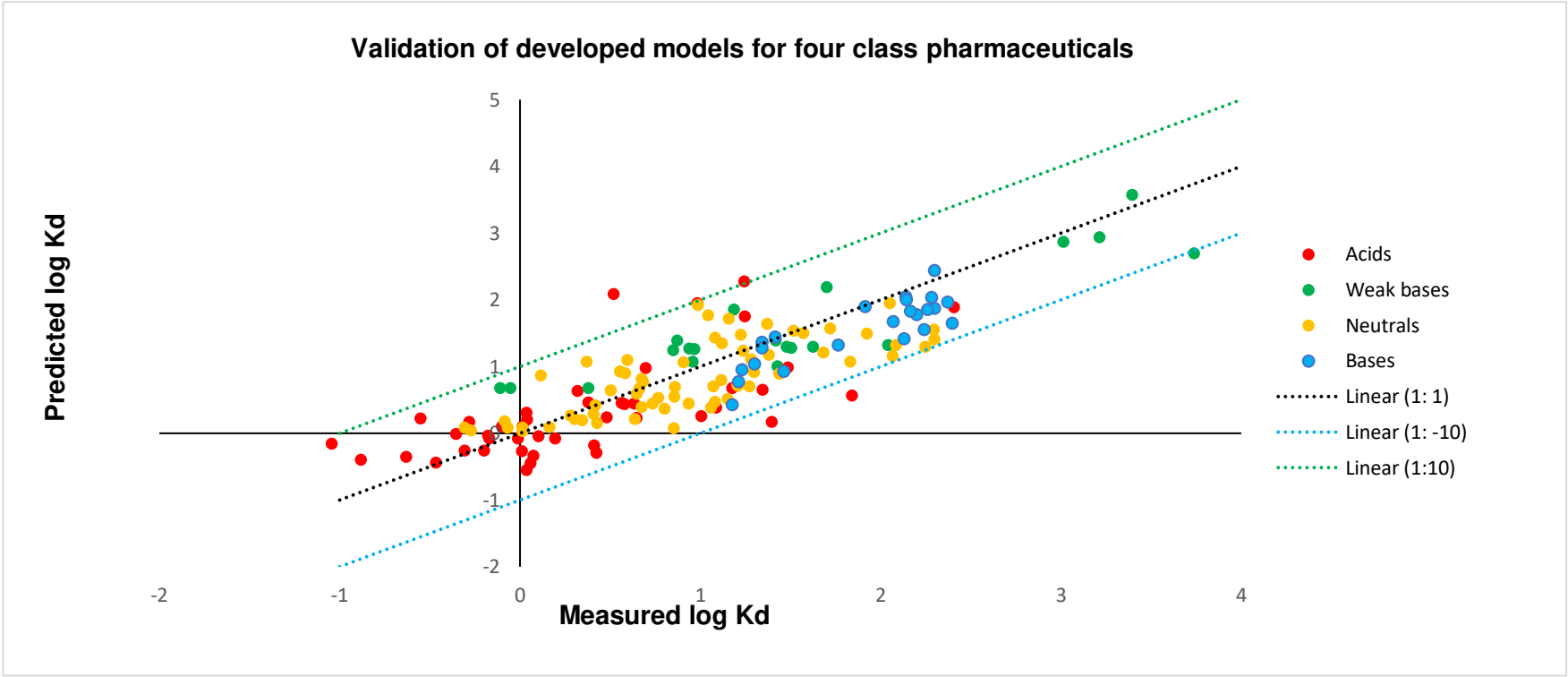
Evaluation data set	N	SD	Existing model	R^2_{test}	$RMSD_{test}$	NSE
Neutral	65	0.637	Franco and Trapp (2008)	0.521	0.601	0.096
Bases	23	0.447	Franco and Trapp (2008) base model A	0.789	0.417	0.088
			Franco and Trapp (2008) base model B	0.628	0.647	-1.194
Weak bases	20	1.082	Franco and Trapp (2008) base model A	0.512	0.903	0.267
			Franco and Trapp (2008) base model B	0.504	0.811	0.409
Acids	44	0.733	Franco and Trapp (2008)	0.547	0.573	0.375
			Franco et al. (2009)	0.513	0.558	0.406
			Kah and Brown (2007)	0.499	0.870	-0.441
			European Union (2003).	0.348	0.611	0.288

N is the number of the observations.

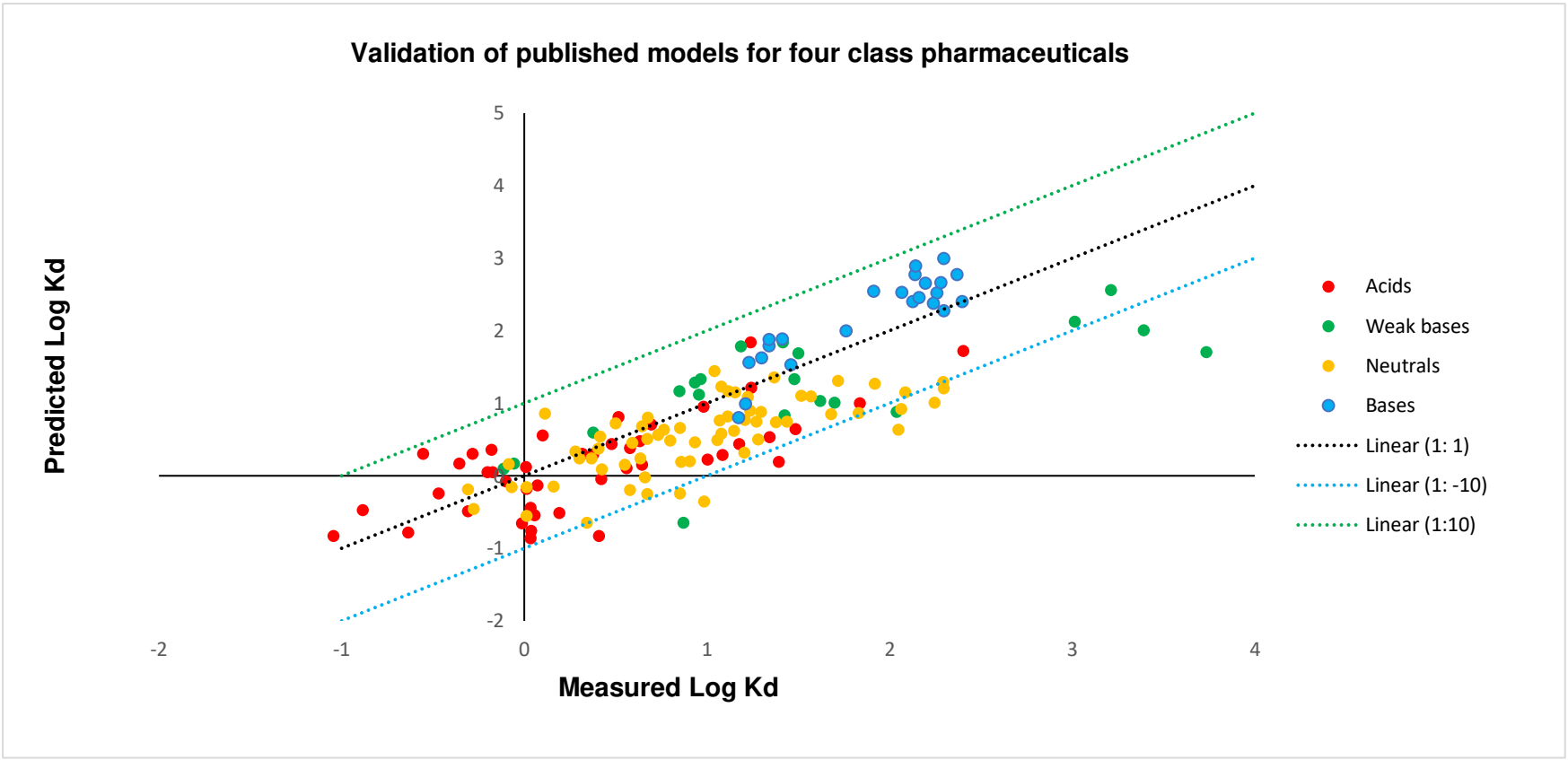
SD is the standard deviation of the observations.

$RMSD_{test}$ is the root mean square deviation.

NSE is the Nash−Sutcliffe Efficiency value.


Validation of developed models for four class pharmaceuticals
Predicted log Kd
Measured log Kd
5
4
3
2
1
0
-1
-2
-2
-1
0
1
2
3
4
Acids
Weak bases
Neutrals
Bases
Linear (1: 1)
Linear (1: -10)
Linear (1:10)

(A)



(B)

Figure 3. Comparison of predictive performance between the developed models in the current study and existing models in the literature. The selected models for the comparison were the model showing the best performance in each class (The model performance results are presented in Table 3 and 4). A) Validation of models 1, 3, 4, 6 developed in present study for neutrals (Log Kow > 0.85), bases (*pKa* > 8), weak bases (8 > *pKa* > 4.8), acids (6.8 > *pKa* > 3.2), respectively; B) Validation of the existing models for bases, weak bases and neutrals proposed by Franco and Trapp [27] and the model for acids proposed by Franco et al. [26]. The black dashed line represents perfect model fit (1:1 line) and the green and blue dashed lines represent a difference of 1 order of magnitude.